

From sensor data to the perception of phenomena: Software architecture for online access and offline analysis

Nader Cheaib, Anne Ruas, Olivier Gaborit

French Institute of Science and Technology for Transport, Development and Networks (IFSTTAR), *Cité Descartes - 14-20 bd Newton – 77447 Champs-sur-Marne – Marne-la-Vallée, France*

Abstract. In this paper, we present a software architecture that enhances the 3D visualization of sensors' data. We aim to better explore and analyze sensors' data dealing with atmospheric pollutants. Hence, our software architecture provides two techniques to interact with our geospatial data : a light web application for an easy access and visualization of this data, and a 3D GIS that allows an enhanced analysis of the geospatial data as well as the atmospheric pollutants. In consequence, this work helps at proposing better civil and urban engineering strategies in order to reduce pollution rates, and hence, enhance urban life quality.

Keywords: 3D visualization, sensors' data, 3D GIS

1. Introduction

In today's societies, we are witnessing an increasing shift of habitations from rural to urban areas, and more specifically towards big cities. This is naturally due to the fact that people tend to live closer to their work, which generates urban sprawl that occupies more and more land and eventually arise commuting time. To fight against these effects, people want to live more in cities, which arises density. The society is encouraging people's concentration in order to reduce commuting time, and therefore reducing energy bill, stress and psychological concerns due to that matter, while at the same time preserving rural environments. However, with the increase of habitations in and around urban areas, people are in increasing need for more comfort in their homes as well as in their everyday life. As urban areas become denser, specific needs arise in order to bring more comfort to people that are looking for green areas and low pollution. Hence, finding

adequate solutions to these constraints is a major issue that should be well taken care of. Moreover, sustainable development implies a better preservation of patrimony and human-made structures for continuous occupancy. For this reason, people are more motivated than ever to handle the consequences of natural phenomena, as well as the consequences of environmental parameters, such as atmospheric pollutants and buildings thermal measurements. These aspects play a major role in enhancing life quality in urban environments. However, measuring and capturing this form of continuous data can be a burden for technical teams to remedy these problems. Since many years now, installing sensors in critical geographic areas has been widely approved as a successful way to capture data in real time.

This paper focuses on the components and functionalities needed to facilitate the usage of sensors' data by numerical models. Our system deals with atmospheric pollutants, as well as thermal measurements in buildings' walls, where experiments are performed in an urban area in Paris' suburbs in the context of the Immanent project. Hence, we propose a software architecture that enhances the analysis of numerical models' results, using either an online access through web service technologies or through an enhanced (offline) analysis using a 3D GIS. In what follows, we present some related work in the domain of sensors' data visualization and analysis. Then, we propose a software architecture that aims at handling data flow between the various components that build our system. Afterwards, we propose a UML study for an enhanced 3D visualization and analysis of sensors' data. Then, we discuss the visualization and analysis process of atmospheric pollutants using a 3D GIS. We finish by a conclusion and future directions in the field.

2. Related Work

Sensor-based systems are increasingly used to extract data in various and wide range of domains, such as for example traffic surveillance and control of vehicles' flow, but also to analyze data related to weather forecast and natural phenomena to prevent from natural disasters. (Bouju 2009, Richeter 2009) define the various types of sensors. In (Bouju 2009), the authors classify sensors into three main types: fixed, mobile and agile. The fixed sensors have a fixed geographic location, where only the values captured by the sensor evolve over time. Mobile sensors are characterized by a change in their position that is continuously evolving, where this position is often the most important data to analyze. Finally, agile sensors "discretely" change their location between two measurement extractions of the same phenomenon. These sensors combine the features of both fixed and mobile sensors, while managing to stay fixed for a long period of time,

and then can become mobile for short periods. In our work, we assume that the sensors' locations are fixed, but they might be agile as well. What is important to our work is the information extracted and related to the phenomena observed. (Servigne 2011) proposes a UML study for a network of sensors that describes the various attributes related to the information extracted (values, precisions etc.), position coordinates, as well as information related to the sensor itself identified by a unique "id", with its autonomy and memory capabilities. This study helped in handling spatio-temporal meta-data related to the network of sensors' that is agile or mobile in nature. (Servigne 2011) finally mentions that the visualizing process of sensors' data must be tailored according to the user: an interface showing road traffic for a politician to make a decision for building a new bridge should be different from the information represented to a traffic controller for synchronizing red lights'.

(Lange 2011) proposes an approach to visualize sensors' data related to energy consumption in a green data center in IBM Montpellier (France). The authors propose a 3D visualization and interaction technique based on a "particle" approach using a segmentation algorithm based on Voronoi cell extraction for space decomposition in 2D or 3D. The aim is to detect a collision between sphere and point in order to find the nearest sensors for each particle. A second technique is the use of Delaunay triangulation that gives a better approximation to particles weight, so that their locations can be determined according to the mesh of sensors. (Fan 2004) explores an approach to process and interpret sensors' data to monitor plants or other endangered species. This approach uses database technologies and web interfaces to visualize the collected data. An interpretation system has also been put in place in order to convert raw climate data into user-friendly visual formats. In this approach, the authors use an algorithm for generating a Voronoi diagram to divide a 2D map into different portions. The visualization process is based on specific shapes and colors to visualize different attributes of sensors' data. Finally, the authors use a 2D GIS (GRASS) in order to visualize this data in the area of the University of Hawaii at Manoa.

(Rodriguez et al. 2012) present a prototype application in order to visualize sensor data retrieved from NASA SOS servers on a virtual globe. The authors affirm that virtual globes provide means to perceive 3D aspects of geographical features, and sensors are one of the data sources that can be used to represent virtual globes. The authors argue that a data taxonomy is needed to analyze the best visualization method, based on cardinality, nature and dimensions of the data in order to describe its temporal and spatial behavior. They have implemented a prototype called World Wind Java (WWJ) that allows the creation of several types of annotation, shapes

lines and custom rendered elements. Finally, the visualization process provides means to configure the way visual aspects of visualization are related to the data itself, where the basic data visualization types proposed are line and time series charts, as well as scatter plot charts in a 3D scene.

(Gutierrez-Rodriguez et al, 2007) argue that there are still many issues to solve concerning the real-time management of spatio-temporal data coming for sensors. The issues concern specifically creating a data structure in order to facilitate their extraction and their smart communication to users. The author proposes a UML conceptual model to evaluate the metadata quality of real-time sensor data that measure natural phenomena. The model proposed is able to evaluate the relation between spatio-temporal data and real-time data, while facilitating the distinction and the identification of the metadata from the data itself. It also acts as a support for data evolution in their context of use. However, a visualization scenario of sensors' data in this model is still not proposed.

(Rude and Beard, 2012) explore a specific problem in interpreting sensor data, which is the detection of high-level events from spatially distributed time series. For the authors, a high-level event refers to a complex, multivariate phenomenon with spatial properties that is evolving over time (storm, forest fire, etc.). Hence, their goal is to extract interesting occurrences from time series of these types of events as aggregate patterns of primitive patterns. By primitive events, the authors refer to a subsequence of a time series for which a particular property of a parameter holds, indicating a state or change of state over a temporal interval. In this context, (Peuquet, 2012) affirms that temporal and spatio-temporal patterns of multiple event types tend to have hierarchical structures having complex interactions and feedback effects. This constraint makes it hard to understand past and on-going real-world events, which require meaningful patterns that may not be visible within complex real-world data. The author proposes a statistical temporal pattern discovery technique called T-pattern analysis via a prototype called Stempo, which is designed to detect patterns of associations among event-types from data organized as series of timelines. Each timeline consists of time series when events of a single event type occurred. This technique has been proven to provide a better understanding of complex space-time interactions within event data.

In what follows, we propose a software architecture to visualize and exploit sensors' data through a web application and a 3D GIS for an enhanced analysis. We do not focus our study on the connection between sensors such as (Bröring et al, 2011) but on the exploration of phenomena.

3. Software Architecture for online access and offline analysis of sensor data

Figure 1 shows our software architecture as well as the data flow in the system. In our work, we use fixed sensors that contain electronic components to process and send data through wifi connexion. We call them Sensorbox as they contain many sensors of different nature. We give a brief overview of each component in the system.

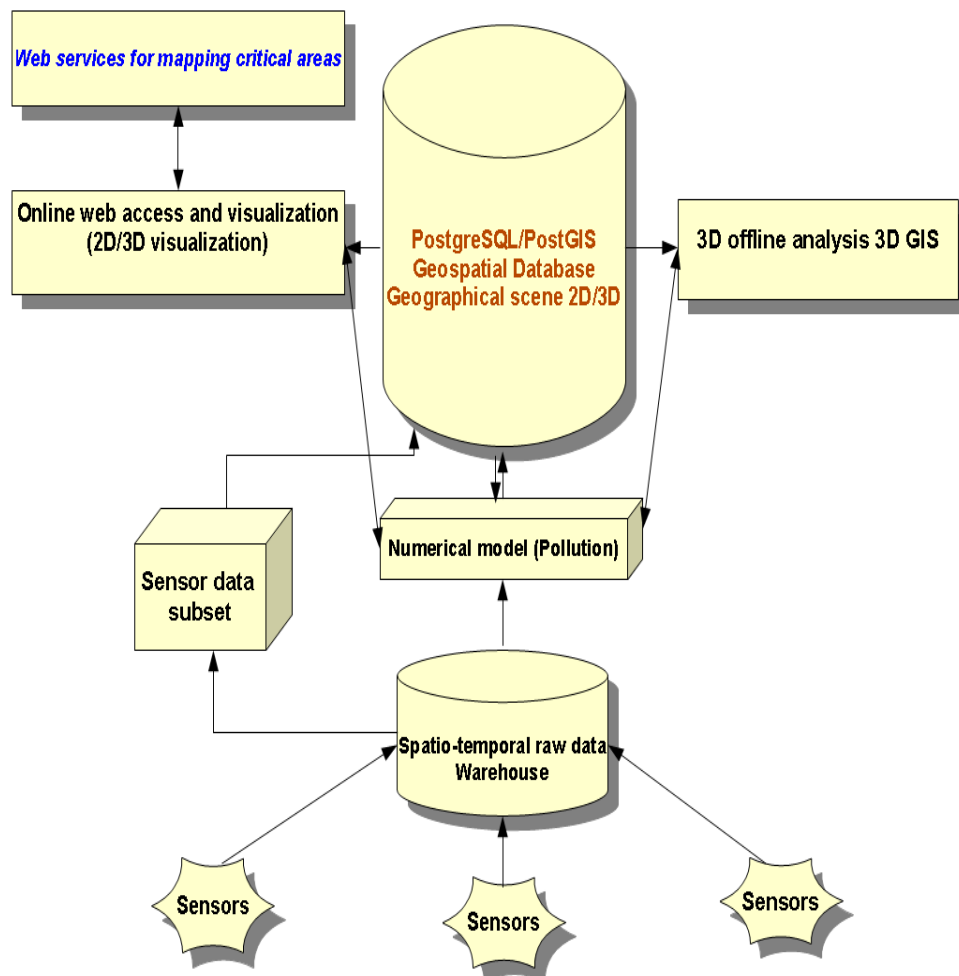


Figure 1. The global software architecture of the system

3.1. PostgreSQL/PostGIS

We use PostgreSQL, which is an open source object-relational database system that insures interoperability between heterogeneous applications. However, as our software architecture eventually require storing 3D spatial objects (such as buildings), we have used PostGIS for this purpose. In fact, the DBMS is a key component as we want to integrate new attributes based on sensors' data directly in the tables where 3D objects are stored. We are actually able to store 3D structures into PostGIS through converting them from **Solid** types into **MultiSurface** or **MultiPolygon**. This is due to the fact that PostGIS is still not able to recognize objects with their geometries defined as **Solid**.

3.2. Numerical Model – Pollution

A numerical model has been put in place based on the code Saturn (open source) platform, which is the EDF's (Electricity of France Company) general purpose Computational Fluid Dynamics (CFD) software. This model takes as input a subset of sensors' data along with the geographic scene, and gives the atmospheric pollutants level in this geographic location. In our work, we are solely interested in the input/output data and data formats that the numerical model (considered as a black box) handles, as we want to connect it with other components in the system to insure a reasonable data flow. Hence, the objective is to automate the execution of this numerical model from the web interface by the user himself. Pollution models are computed by the CEREAs (<http://cerea.enpc.fr>) (see figure 5)

3.3. Online Access and Visualization

The nature of geographical applications requires seamless integration of spatial data from a variety of providers. This constraint motivated researchers to deploy new technologies that offer the desired interoperability to access and visualize geospatial data on the web, which are geographic web services. Two other standard protocols are crucial in building a geospatial web service in our system: Web Map Service (WMS) and Web Feature Service (WFS). WMS produces maps from geographic data through regular GET and POST HTTP request methods. WFS specifications store and serve the geospatial data encoded in Geographic Markup Language (GML), which is an XML encoding for the transport and storage for geographic information including both for spatial (attributes) and non-spatial (geometric) properties of geographic features.

3.4. 3D GIS for offline data analysis

We use the GeOxygene (Brasebin 2009) software for our offline analysis and exploitation of sensors' data. Geoxygene is an open source framework that was initially built for the deployment and visualization of 2D GIS applications, but was later extended to 3D visualization. It is developed in

Java using ISO standards and OGC specifications. Figure 2 shows a snapshot of the Geoxygene interface that enables 3D visualization of geographic scenes. We have created a connexion between Geoxygene and PostgreSQL/PostGIS to better manipulate 3D objects. Hence, the DBMS serves as the main component linking the numerical models with Geoxygene in order to visualize natural phenomena in a 3D scene.

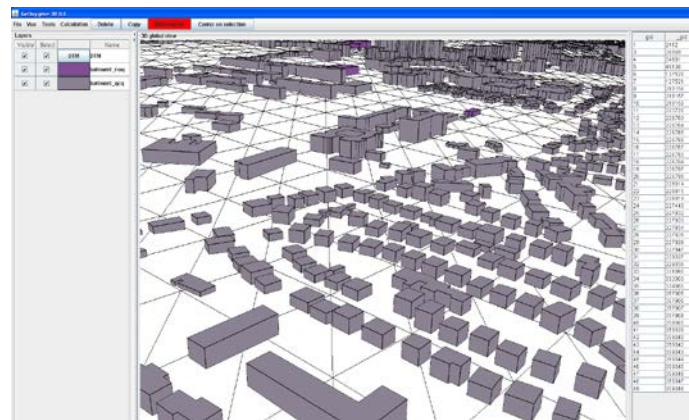


Figure 2. Screenshot of Geoxygene for 3D visualization of geographic scenes

The task of extracting knowledge about geospatial processes from spatio-temporal datasets are quite complicated, as these datasets are often large, highly-linked, complex and are not very accurate. Hence, researchers in the field must adapt existing algorithms and computational technologies in order to better exploit this “big data”. At this stage of our work, we are not specifically dealing with enormous data quantities, however, the data extracted from the sensors are big and unstructured enough in order to implement a special component, represented by a cube in Figure 1 (on the left), which will extract and structure the most important data from the data warehouse, and inject it in the spatial DBMS along with the data output coming from the numerical models. This component insures data interoperability between the data warehouse and the spatial DBMS containing the geographic data to be integrated into the 3D GIS.

3.5. UML Design

Figure 3 shows the various classes in our system, as well as the data types used. The first component is the Sensorbox that extracts information related to data's localization, nature, size as well as an identifier for the buildings' wall in case it is related to thermal measurements. It is then sent through a Wifi connexion to a data warehouse, which by default saves all the incoming data without any filtering process. After this step, the data can

be injected into two types of numerical models: The first model uses this data in order to produce properties concerning atmospheric pollutants in the area of study. The second model uses this data to generate thermal measurements concerning building's walls. In this paper, we concentrate our effort on the visualization of atmospheric pollutants. The final phase is to integrate the geographic scene studied with the data from the numerical models. The result would be an enriched 3D geographic scene containing the values of atmospheric pollutants. In Figure 4, we can see the data flow process

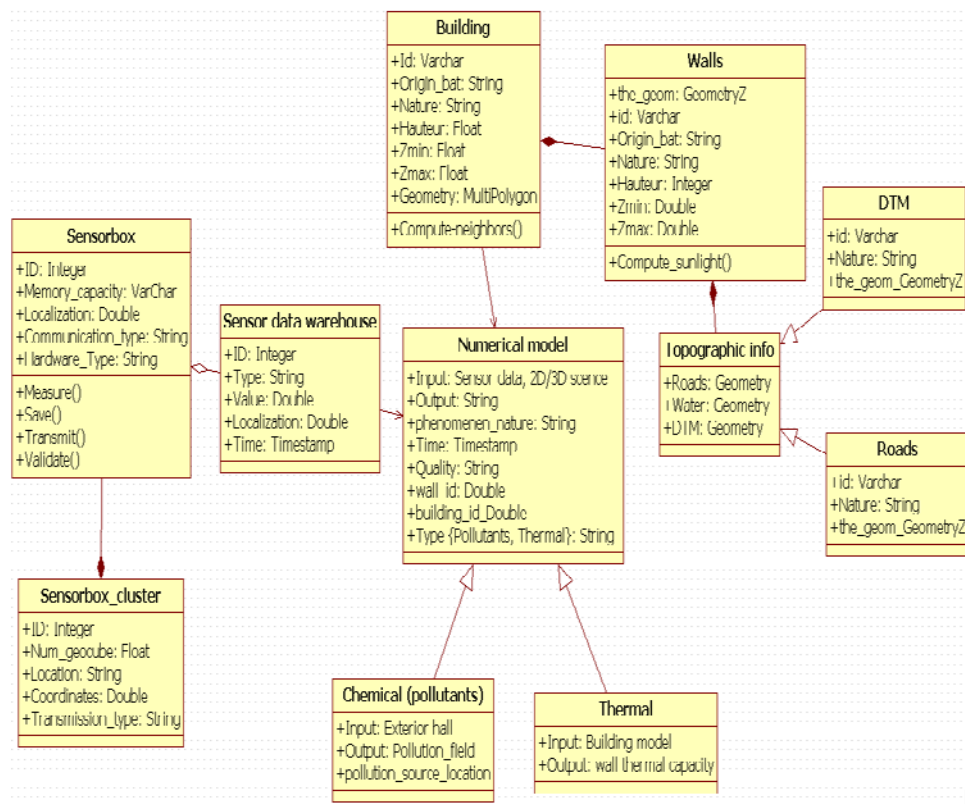


Figure 3. UML Design of the general architecture of our system

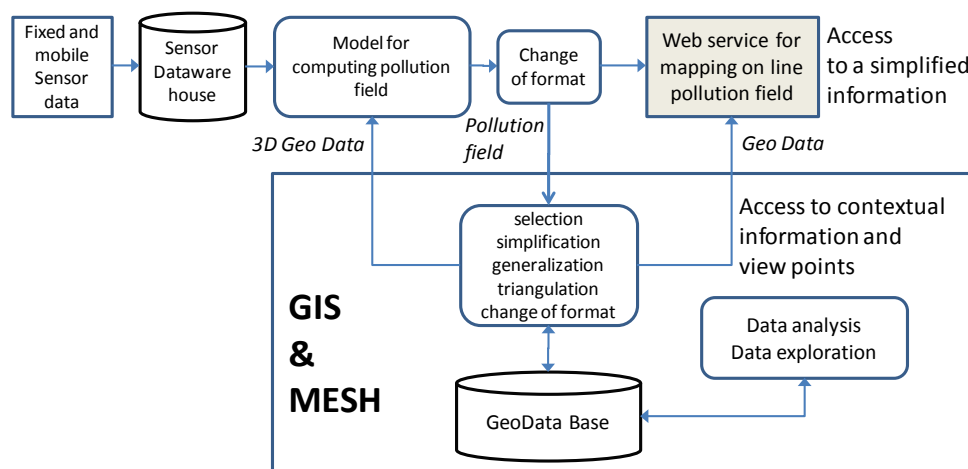


Figure 4. The role of GIS in the data flow from sensors to web viewer and data analysis.

An easy way today to view pollution is to create a flat grid from the pollution model and to view it as if it would be a DTM (see figure 5). This information can easily be mapped but it is flat and not appropriate for analysis.

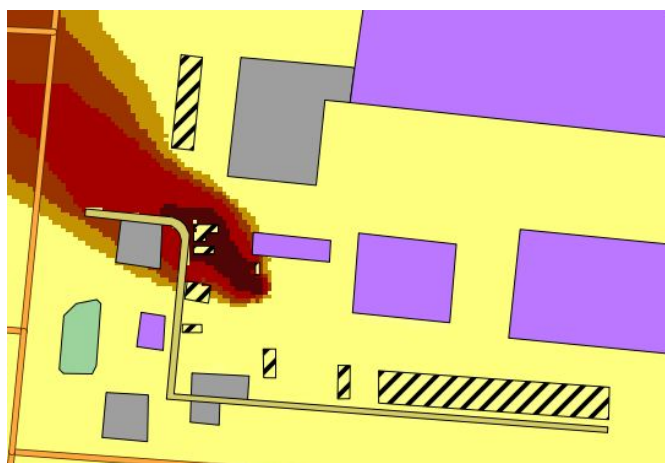


Figure 5. Pollution can be mapped as a DTM (Boukhechba 2013)

4. Towards a GIS Prototype to Map Environmental Data on 3D Scene

In this section, we define the conceptual schema and some first graphical proposals to map environmental data on a 3D GIS. We want to conceive a platform to explore phenomenon from different viewpoints, while studying their co-existence in a geographical scene. Our assumption is that geographical information should not be viewed only as *ornament* data,

'data to make nice graphics', but that it plays an important role for a better understanding of phenomena. Not only space plays an active role in the computation of phenomena (e.g. to compute wind flow or sunlight exposure) but geographical data describes the places where humans live: work, sleep, eat, commute or travel. As an example, (Valente et al 2012) show a computation of particle matter exposure around and in a school. This research illustrates the use of geographical data to compute not only the dispersion of pollutants but also the exposure of children to these pollutants in specific and exposed places.

We believe that GISs constitute an excellent basis to explore phenomena as GISs are meant to show and analyze the co-existence of things. If physics is excellent to model phenomena, geography has the appropriate paradigms to show and analyze what is happening here? (Why?) What are the effects of these co-occurring phenomena here? In this subsection, we propose to add information to the GIS to facilitate data exploration.

4.1. Data modeling for visual exploration

Let us consider a phenomena described by a value field (such as temperature, pollutants) and different time series. Of course, this information is too large to be viewed. If we consider this data, we may need to reduce the number of data by aggregating values (Spatial average pollution field) and even plans that represent the data on a specific surface (often vertical or horizontal plan, but it could be a part of a sphere). A surface offers the possibility to view a phenomenon from a specific view point and direction. It could be the pollution at the specific altitude or the noise from a specific view point. Then, the value could be either represented on points according to the granularity of the required grid, or it can be aggregated to create areas regrouping values in a specific range of values. These graphic representations are common in physical viewer software such as ParaView. In our work, we propose to explicitly add these classes of information and related methods to compute these values on demand for exploration purposes. These views can be built for a specific time but it is also interesting to study the phenomena for a longer period of time. For example, to study if a phenomenon depends on seasons, diurnal rhythm or other criteria as well. It is thus necessary to use methods to aggregate data not only on spatial scale but also on a temporal one. In Figure 6 below, we can see a data schema for representing pollution data using a GIS

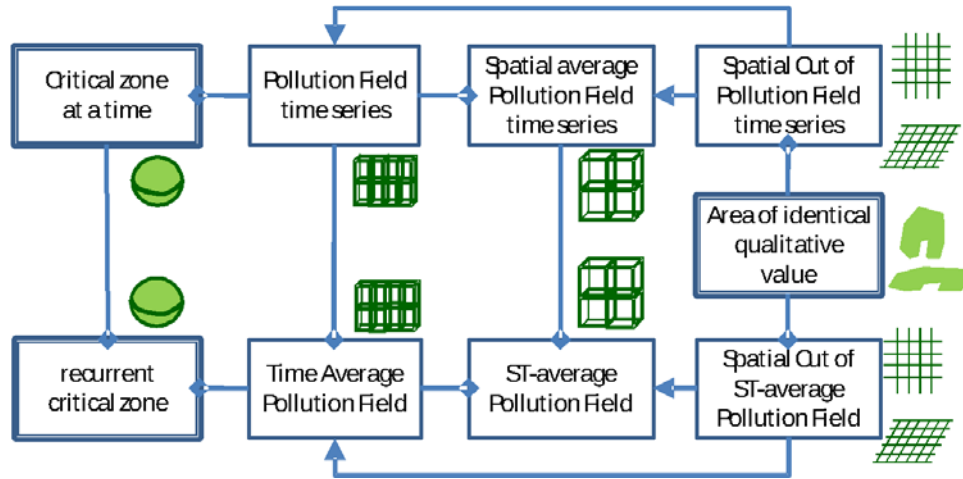


Figure 6. Data Schema for representing pollution data on a GIS

Moreover we introduce the concept of **critical zone**, which is a geographical location (a volume) that has characteristic values such as extreme ones or rare ones. These *critical zones* are detected by means of classical spatial analysis methods. They can be detected for a specific time or during a time frame. In such a case, we call them **recurrent critical zones** that represent redundant local phenomena (such as a part of a street that is often noisy or polluted or both). Critical zones can actually be locations that cumulate different rare or unpleasant phenomena. We call them “critical” because we want to evaluate the values with a specific viewpoint.

4.2. First developments

The data schema in Figure 6 requires the development of specific geometric methods. We made first experiments on GeOxygène 3D platform. We used real topographic data from the IGN-France BDTopo © and added some sensors as specific objects, and fictive sensor values.

The first development consists in defining a volume, which is a location where the visual analysis will be done. As we know that the field of values needs a lot of computational time and a big storage capacity, we want to be able to define ‘working space’ where the analysis is performed (Figure 7 left). Then, we can find all objects inside the volume and the ones that are cutting it. We also defined the methods to build a plan. These two methods are tuned with several parameters that the user can manipulate using a dedicated interface: a center point, a distance and a direction.

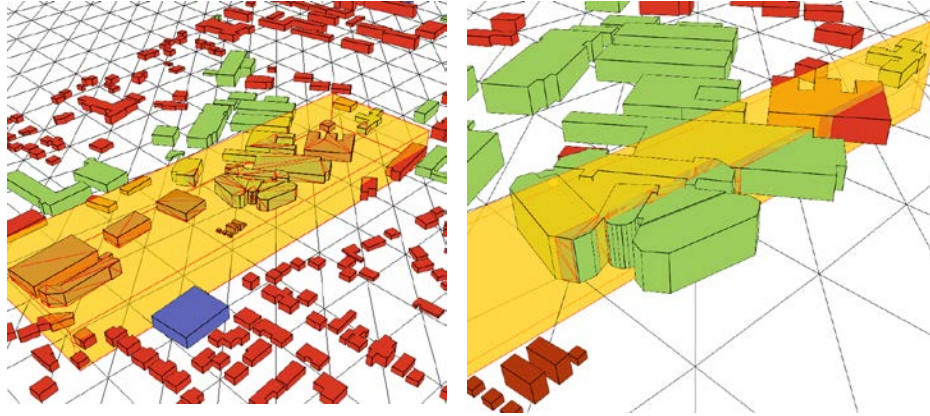


Figure 7. Reducing the space (left) or computing a plan (right) © IGN-BDTopo

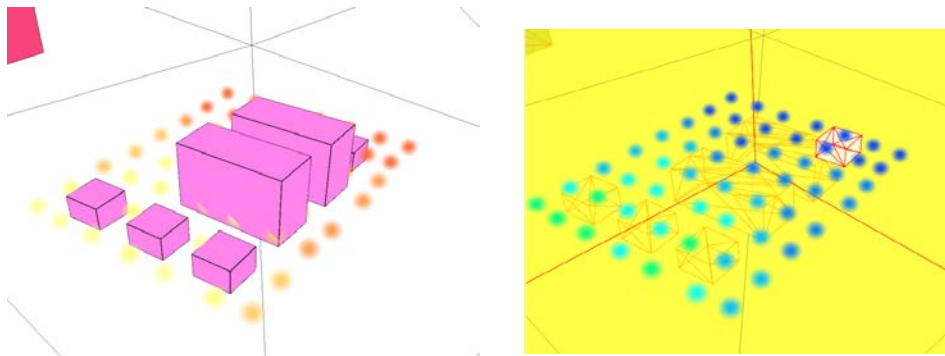


Figure 8. Visualization of pollutants as billiard balls or cloud effect

Then, we developed methods to map a set of values, on a cut (a plan in Figure 8) with different graphic solution such as billiard balls or cloud effect. For experimental purposes, we created a grid of points on a plan previously set (cut), and we mapped these values on the chosen plan or inside the volume. As values vary between 0 and 10, we defined associated symbols as shown in Figure 8.

5. Conclusion and Perspectives

In this paper, we presented a software architecture that enhances the 3D visualization of sensors' data related atmospheric pollutants. We presented a first version of a software architecture that enables the visualization and analysis of sensor data: 1) through a web application for an easy access to geospatial data by the general public, and 2) using a 3D GIS for an (offline) enhanced analysis of initial data coming from sensors. We do believe that

our work constitutes a first step towards visualizing atmospheric pollutants using an online and an offline access and analysis techniques, which ultimately help in proposing recommendations for future urban developments.

References

- Boukhechba M. (2013) Mise en œuvre d'une architecture logicielle à base de Services web pour la visualisation de données de pollution dans une scène géographique 2D en ligne. Rapport de stage Master Géomatique 2013.
- Bouju A, Bertrand F, Servigne S, Devogèle T, Ray C, Martin H (2009). Gestion de données spatio-temporelles au sein de base de données capteurs. In the COGIST'09, pp. 164-172. ST-Quay-Portrieux, France.
- Brasbin M (2009) GeOxylene: An Open 3D Framework for the Development of Geographic Applications. Proc of 12th AGILE, Germany.
- Bröring, A., F. Bache, T. Bartoschek & C.P.J.M. van Elzakker (2011): The SID Creator: A Visual Approach for Integrating Sensors with the Sensor Web. In: S. Geertmann, W. Reinhardt & F. Toppen (Eds.), Advancing Geoinformation Science for a Changing World, Lecture Notes in Geoinformation and Cartography, Springer, Volume 1, pp 143-162.
- Fan F, Biagioni E (2004). An approach to Data Visualization and Interpretation for Sensor Networks. In the 37th Hawaii International Conference on System Sciences.
- Gutierrez Rodriguez C., Servigne S. and Laurini, R. (2007) Towards real time metadata for network-based geographic databases. proceedings of ISSDQ2007, 5th International Symposium of Spatial Data Quality, pp. 13-15.
- Lange B, Rodriguez N, Puech W, Rey H, Vasques X (2011). 3D Visualization of particle system with extracted data from sensor. In the 9th LACCEI Latin American and Caribbean Conference, Medellin, Colombia.
- Peuquet, D.J (2012). A Method for Discovery and Analysis of Temporal Patterns in Complex Space-Time Event data. In the Extended abstracts of the 7th International GIScience 2012 Conference, Springer Verlag.
- Rodríguez L.E, Tamayo A, Beltrán A and Huerta J (2012). Visualization of Sensor Data in Virtual Globes. Proceedings of the AGILE'2012, pp. 41-45, Aignon, France.
- Rude A, Beard K (2012). High-Level Event Detection in Spatially Distributed Time Series. In the proceedings of the 7th International GIScience 2012 Conference, pp. 160-172, Springer Verlag.
- Servigne S, Gutiérrez C (2009). Métadonnées et qualité pour les systèmes de surveillance en temps réel. In the International journal of Geomatics, V 19, N 2, pp. 151-168.

Valente J, Amorim JH, Cascao P, Rodrigues V, Borrego C (2012) Children exposure to PM levels in a typical school morning. In Usage, Usability and utility of 3D City models Conference, Nantes, France.